

Genetic Bit Analysis: a solid phase method for typing single nucleotide polymorphisms

Theo T.Nikiforov, Robert B.Rendle, Philip Goelet, Yu-Hui Rogers, Michael L.Kotewicz, Stephen Anderson¹, George L.Trainor² and Michael R.Knapp*

Molecular Tool, Inc., Alpha Center, Hopkins Bayview Research Campus, 5210 Eastern Avenue, Baltimore, MD 21224, ¹Center for Advanced Biotechnology and Medicine, 679 Hoes Lane, Piscataway, NJ 08854 and ²Chemical and Physical Sciences, Research and Development, Du Pont Merck Pharmaceutical Company, PO Box 80353, Wilmington, DE 19880, USA

Received June 29, 1994; Revised and Accepted September 7, 1994

ABSTRACT

A new method for typing single nucleotide polymorphisms in DNA is described. In this method, specific fragments of genomic DNA containing the polymorphic site(s) are first amplified by the polymerase chain reaction (PCR) using one regular and one phosphorothioate-modified primer. The double-stranded PCR product is rendered single-stranded by treatment with the enzyme T7 gene 6 exonuclease, and captured onto individual wells of a 96 well polystyrene plate by hybridization to an immobilized oligonucleotide primer. This primer is designed to hybridize to the single-stranded target DNA immediately adjacent from the polymorphic site of interest. Using the Klenow fragment of *E.coli* DNA polymerase I or the modified T7 DNA polymerase (Sequenase), the 3' end of the capture oligonucleotide is extended by one base using a mixture of one biotin-labeled, one fluorescein-labeled, and two unlabeled dideoxynucleoside triphosphates. Antibody conjugates of alkaline phosphatase and horseradish peroxidase are then used to determine the nature of the extended base in an ELISA format. This paper describes biochemical features of this method in detail. A semi-automated version of the method, which we call Genetic Bit Analysis (GBA), is being used on a large scale for the parentage verification of thoroughbred horses using a predetermined set of 26 diallelic polymorphisms in the equine genome.

INTRODUCTION

Mammalian genomes carry numerous single nucleotide polymorphisms (SNPs). On average, two to three polymorphic sites are found per kilobasepair in human genomic DNA (1). Most of these polymorphisms are 'silent' and do not give rise to detectable phenotypes, but an important subset of mutations are associated with heritable diseases such as cystic fibrosis (2),

sickle cell anemia (3), colorectal cancer (4), and retinitis pigmentosa (5, 6).

The wealth of genetic information associated with SNPs can be exploited in a wide variety of applications ranging from the detection of alleles linked to common genetic diseases, to the identification of individuals, to the use of genetic polymorphisms in gene mapping projects. Each of these applications involves the analysis of a large number of samples and will ultimately require rapid, inexpensive, and highly automated methods for typing DNA sequence variants.

Because of the importance of SNPs, a number of methods have been described for their *detection* and *typing*. In general, methods that can be used to discover new mutations can be applied to the typing of those that are already known. These methods include restriction fragment length polymorphism (RFLP) analysis (7), denaturing gradient gel electrophoresis (8), single strand conformation polymorphism (SSCP) detection (9), and chemical or enzymatic mismatch modification assays (10,11). Although powerful, these techniques typically rely on electrophoretic separation to detect the polymorphisms, are relatively labor-intensive, and are difficult to automate.

Approaches for the large-scale typing of known mutations have been described that can be carried out in a nonelectrophoretic mode. Some of these approaches rely on differential hybridization to discriminate the different alleles (12, 13). Other methods, which base the discrimination on an enzymatic reaction, include the oligonucleotide ligation assay (OLA) (14,15), the ligase chain reaction (16), the allele-specific polymerase chain reaction (17,18), and the primer guided nucleotide incorporation assays (19–23). Of these, both the oligonucleotide ligation assay and the primer guided incorporation techniques have been developed to a stage where they can be used for the typing of a relatively large number of samples.

Here, we present a new primer guided genotyping method (Figure 1). The sequence information surrounding the site of variation in the target DNA is used to design an oligonucleotide

*To whom correspondence should be addressed

primer that is complementary to the region immediately adjacent to, but not including, the variable nucleotide site in the target DNA. A single-stranded nucleic acid target molecule is hybridized to this primer immobilized on the polystyrene of a 96-well microplate. The primer is extended by one haptenated dideoxynucleoside triphosphate using a DNA polymerase in the presence of all four chain terminating dideoxynucleoside triphosphates. Novel haptenated ddNTPs allow discrimination of the incorporated nucleotide to be accomplished using standard, enzyme-linked colorimetry.

Our method differs significantly from other primer guided genotyping methods described in the literature. Most importantly, in our method the extension step is carried out in the presence of chain terminating ddNTPs only, and therefore only one nucleotide can be incorporated at the 3' end of the immobilized primer. Second, the immobilization of the primer rather than the template permits the removal of the latter from the reaction mixture following the extension of the primer and thus elimination of all signals that could arise from nonspecific extension at the 3' end of the template. Thirdly, the technique allows the detection of two possible alleles in the same well of a microtiter plate which results in both operational and biochemical advantages.

In this paper, we give a detailed description of this DNA typing method, called Genetic Bit Analysis (GBA). The 'genetic bit' is the term we have adopted for the most elementary form of genetic information, namely a single DNA nucleotide. GBA is a highly flexible method that can be applied, under a standard set of biochemical conditions, to the typing of any nucleic acid polymorphism whose sequence is known. In this paper we focus on the biochemical basis of GBA. Our experience would suggest that features of specificity and convenience inherent in the GBA biochemistry permit the method to become widely used for typing single nucleotide polymorphisms (SNPs) in both research and clinical laboratory applications.

EXPERIMENTAL

Enzymes

Taq DNA polymerase was obtained from Perkin-Elmer. *E. coli* DNA polymerase, Klenow fragment, and T7 gene 6 exonuclease

were purified from recombinant *E. coli* clones containing suitable expression plasmids (unpublished).

Oligonucleotide synthesis

All oligonucleotides were synthesized using standard phosphoramidite chemistry on an Applied Biosystems 392/394 DNA synthesizer, using reagents obtained from Glen Research (Sterling, VA). For the synthesis of phosphorothioate primers, the sulfurizing reagent tetraethylthiuram disulfide (TETD, Applied Biosystems) was used as recommended by the manufacturer. All oligonucleotides were deprotected with concentrated ammonia and desalted using NAP 5 (0.2 μ mol scale synthesis) or NAP 25 (1 μ mol) gel filtration columns (Pharmacia). Oligonucleotides biotinylated at the 5' end were prepared using a biotin phosphoramidite (DMT-Biotin-C6-PA), obtained from Cambridge Research Biochemicals, Inc. (Wilmington, DE). The coupling time of this phosphoramidite was extended to two minutes. The abasic C₃ linker was introduced using the Spacer Phosphoramidite C₃ (Glen Research).

The sequences of the oligonucleotides used in experiments described in this paper are given in Table 1.

Immobilization of oligonucleotides onto 96-well ELISA plates

Immulon 4 plates (Dynatech Laboratories, Chantilly, VA) were used for all experiments shown. Fifty μ l aliquots of a 0.2 μ M oligonucleotide solution in a freshly prepared 20 mM solution of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, obtained from Sigma) in water were added to individual wells of a 96 well plate and incubated overnight at room temperature. The plates were then washed with a solution of TNTw (10 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween-20). The same procedure was used for other experiments, in which EDC was replaced in the immobilization step by NaCl, tetramethylammonium chloride, cetyltrimethylammonium bromide, and octyldimethylamine hydrochloride.

We have tested a number of different commercially available 96 well plates for their suitability for oligonucleotide immobilization. In general, plates that are described as having a more hydrophilic surface gave good results, whereas those with a hydrophobic surface were found unsuitable. Examples of

Table 1. Sequences of oligonucleotides used in this paper (B denotes a biotin residue; X is a C₃ linker; phosphorothioate bonds are located between the underlined residues)

number	sequence	use
308	5'AGCCTCCGACCGCGTGGTGCCTGGT	GBA primer
308T	5'AGCCTCCTACCGCGTGGTGCCTGGT	GBA primer
308M1	5'AGCCTCCXACCGCGTGGTGCCTGGT	GBA primer
680	5'GAGATGCAGCTCTAAGTGCTGTGGG	GBA primer
680T	5'GAGATTCAGCTCTAAGTGCTGTGGG	GBA primer
1112	5'AGTATAATAATCACAGTATGTTAGC	GBA primer
1676	5'BCCACGGCTAACATACTGTGATTATTACTTAGAT	labeled probe
1464	5'BAATAAGGGGAAACAATTCAGCCCA	GBA primer
501	5'GTTATGGGCTGAATTGTTTCCCTAATTT	synthetic template
713	5'TTCTACATTCTTTCTTGTCTGT	synthetic template
1302	5'GGAGAACAGAAACAAGAAATGAATATGA- ATGTAGAAGCAT	synth. template
1473	5'CCACAACAGAAACAAGAAATGAATATGA- ATGTAGAAGCAT	synth. template
1474	5'AACAGAACAAAGAAATGAATATGAATGT- AGAAGCAT	synth. template
1214	5'ACCTTCAAACTCAACTCAGCTCTT PCR primer	
1215	5'TTTACCAATGAGAAGGACATCTAAG	PCR primer

suitable plates include Immulon 4 (Dynatech); Maxisorp (Nunc); and ImmunoWare plates (Pierce). No attachment could be achieved on Immulon 1 (Dynatech) and Polysorp (Nunc) plates.

DNA isolation and PCR amplification

Horse genomic DNA, isolated from swabs of the nasal mucosa, was the source of DNA in all PCR amplifications. A foam-tipped swab on a six-inch plastic stick was inserted one to two inches into the horse's nostril and immediately immersed in transport buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 5 mM EDTA, 0.5% SDS). The swab remained stored in this solution under ambient conditions until arrival at the laboratory. DNA was isolated from this mixture by treatment with a mixture of guanidine hydrochloride and ethanol and adsorption to glass matrices (e.g., MagicTM resin, obtained from Promega, Madison, WI), followed by recovery in 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA. PCRs were carried out in a total volume of 30–50 μ l. The final concentration of the PCR primers was 0.5 μ M. Following an initial two minute denaturation step at 95°C, thirty-five cycles were carried out, each consisting of denaturation (1 min at 95°C), annealing (2 min at 60°C), and extension (3 minutes at 72°C). *Taq* DNA polymerase was used at a concentration of 0.025 units/ μ l. The final composition of the PCR buffer was: 1.5 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl, pH 8.3, and 170 μ g/ml BSA.

Preparation of single-stranded PCR fragments

Single-stranded DNA was prepared from double-stranded PCR products as described (25). One of the strands was protected from exonuclease hydrolysis by the introduction, during synthesis, of four phosphorothioate groups at the 5' end of one of each pair of the PCR primers. Following the PCR amplification, T7 gene 6 exonuclease, diluted in a buffer containing 50 mM Tris-HCl, pH 7.5, 1 mM DTT, and 100 μ g/ml BSA, was added to a final concentration of 0.5 units/ μ l. Incubation with this enzyme was for one hour at room temperature.

Hybridization of single-stranded PCR fragments to oligonucleotides immobilized onto ELISA plates

After the exonuclease treatment, an equal volume of 3 M NaCl, 20 mM EDTA was added to the reaction mixture and 20 μ l aliquots of the resulting solution transferred to individual wells containing the appropriate immobilized oligonucleotide molecule. Hybridization was carried out for 30 min at room temperature and was followed by washing with TNTw.

Labeled dideoxynucleoside triphosphates

All biotin- and fluorescein-labeled chain-terminating 2',3'-dideoxynucleoside triphosphates used in the single nucleotide extension reaction were purchased from Du Pont NEN, (Wilmington, DE). A selection of labeled ddNTPs are commercially available from that supplier (sold as Renaissance non-radioactive products). These compounds are derivatives of amino-propynyl-substituted 2',3'-dideoxypyrimidines or 2',3'-dideoxy-7-deazapurines. The chemistry of these chain terminators and their use in DNA sequencing have been described (24).

Solid-phase primer extension

Following the hybridization step, 20 μ l of polymerase extension mix was added to each well. The extension mix contained 20 mM Tris-HCl, pH 7.5; 10 mM MgCl₂; 25 mM NaCl; 10 mM MnCl₂; 15 mM sodium isocitrate; 1.5 μ M of two unlabeled

2',3'-dideoxynucleoside triphosphates; 1.5 μ M of one biotin-labeled and 1.5 μ M of one fluorescein-labeled 2',3'-dideoxynucleoside triphosphate; and the Klenow fragment of *E. coli* DNA polymerase I (0.3 units per well). The polymerase was diluted in a buffer containing 10 mM Tris-HCl, pH 7.5, 5 mM DTT, and 0.5 mg/ml BSA. The extension reaction was carried out for 10 min at room temperature. The plates were subsequently washed once with TNTw, once with 0.2 N NaOH, and three additional times with TNTw.

Colorimetric detection of the incorporated nucleotides

After the extension step, the wells were incubated for 30 min at room temperature with 40 μ l of 1% BSA in TNTw containing an alkaline phosphatase conjugate of anti-fluorescein (Bioscience International, Kennebunk, ME) and a horseradish peroxidase-conjugated anti-biotin (Vector Laboratories, Burlingame, CA). The dilution factor was 1:500 for the anti-fluorescein and 1:1000

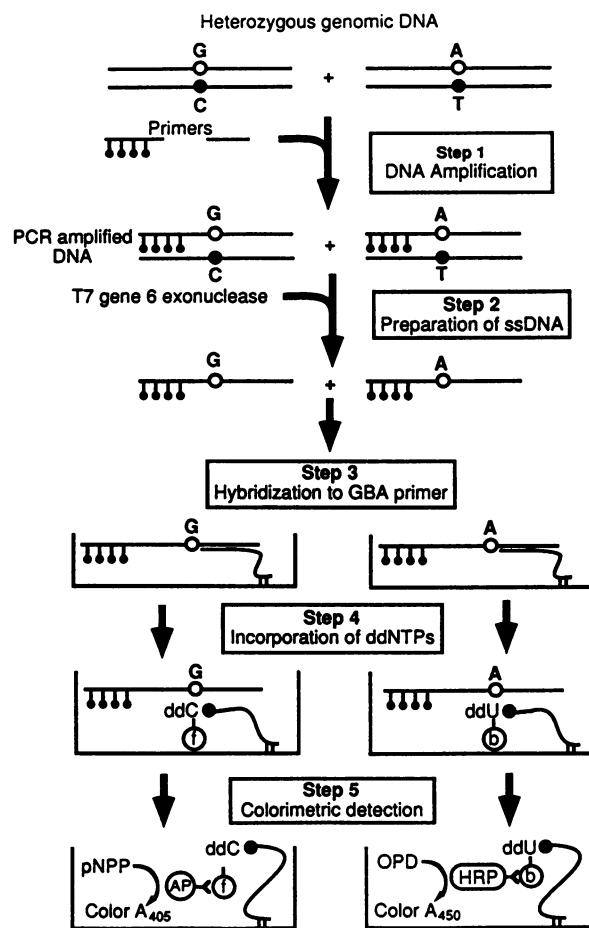


Figure 1. Schematic representation of the individual steps of single nucleotide typing by GBA. In Step 1, a DNA fragment containing the polymorphic site to be typed is amplified by PCR using one primer containing four phosphorothioate bonds at the 5' end. In Step 2, the double-stranded PCR product is rendered single-stranded by treatment with T7 gene 6 exonuclease. In Step 3, the single-stranded DNA template is captured by hybridization to a primer immobilized to the wells of a microtiter plate, whereby the polymorphic site of the template is located immediately downstream from the 3' end of the primer. In Step 4, the 3' end of the primer is enzymatically extended by one nucleotide using haptened ddNTPs. In Step 5, the nature of the incorporated nucleotide(s) is revealed by an enzyme-linked assay.

for the antibiotin. These dilutions were calibrated to give colorimetric signals of approximately equal intensity with the two alleles.

After washing, the presence of alkaline phosphatase was detected first by addition of 100 μ l per well of a 1.5 mg/ml solution of p-nitrophenyl phosphate in 100 mM diethanolamine, pH 9.5, 20 mM $MgCl_2$. The plate was immediately placed in a kinetic microplate reader (Molecular Devices Corporation, Menlo Park, CA), and the development of color was followed at 405 nm for 2 min. The results were expressed as mOD_{405}/min . The plates were then washed again and incubated with 100 μ l of a 1 mg/ml solution of o-phenylenediamine in 0.1 M citric acid, pH 4.5, containing 0.012% H_2O_2 . The reaction was followed by measuring the change of light absorbance at 450 nm as above. Most experiments in microtiter plates described in this article have been carried out at least in triplicate, and the results presented are the averaged numbers. The two enzymes used, alkaline phosphatase and horseradish peroxidase have significantly different pH optima (9.8 vs. 4.5), thus, if endpoint readings are to be taken, it is preferable that the incubations with the two antibody conjugates are carried out sequentially rather than simultaneously in order to avoid partial inactivation of the second antibody conjugate.

Single nucleotide primer extension in solution

A solution containing 2 pmole of the synthetic template # 501 (see Table 1 for oligonucleotide sequences) and 800 fmole of the 5' biotinylated primer # 1464 in 20 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 25 mM NaCl, 10 mM $MnCl_2$ and 15 mM sodium isocitrate, was heated to 95°C for 10 min, then slowly cooled down to room temperature to anneal the primer to the template. Aliquots of this solution were then added to four individual tubes, each containing solutions of one fluorescein-labeled ddNTP, the three other unlabeled ddNTPs, and the Klenow polymerase. After 10 min incubation at room temperature, aliquots of these mixtures were transferred to individual wells of an avidin-coated microtiter plate to capture the extension complexes via the 5' biotin residue of the primer. The wells were then washed with 0.2 N NaOH to remove the template strand and the presence of fluorescein was detected using an anti-fluorescein HRP conjugate.

RESULTS AND DISCUSSION

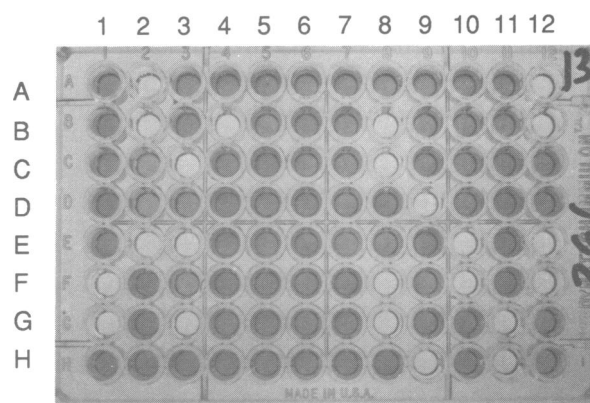
DNA typing by Genetic Bit Analysis (GBA)

The individual steps of GBA are shown schematically in Figure 1. We have developed a test for the parentage verification of thoroughbred horses based on GBA, whereby each horse is typed at 26 different, diallelic loci. The use of 96 well microtiter plates has allowed us to develop a semi-automated version of the test, taking advantage of a number of commercially available, automated liquid handlers for that format. In this automated version of the test, 88 horses are typed, together with suitable controls, with respect to one locus on one microtiter plate. Figure 2 shows a typical result from such a test. This Figure represents the results from the typing of 88 horses with respect to locus JH261-1, a single nucleotide polymorphism present in the equine genome (manuscript in preparation). The same microtiter plate was photographed after development of the colorimetric reaction for alkaline phosphatase which reveals allele 1 (incorporation of fluoresceinated ddCTP, top) and, with appropriate processing,

after the colorimetric reaction for horseradish peroxidase that reveals allele 2 (incorporation of biotinylated ddUTP, bottom). Controls for specific and non-specific effects were also run (see legend to Figure 2). Genotypes are visually scorable: CC homozygotes give a strong reaction with alkaline phosphatase but are negative for horseradish peroxidase, TT homozygotes have the opposite profile, and CT heterozygotes are positive for both enzymatic reactions.

Absorbance values for a comparable set of 88 horses typed with respect to the same polymorphic locus JH261-1 were

A.



B.

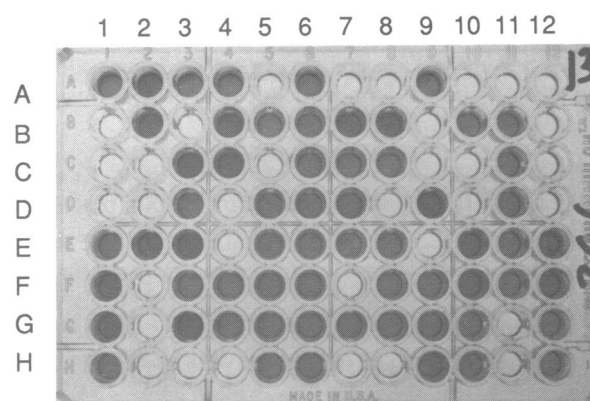


Figure 2. Colorimetric detection of two alleles on a single microtiter plate. The GBA primer # 1112 (see text for sequence) was immobilized in all wells of this plate using EDC. A 116 bp fragment was amplified from the genomic DNA of 86 different thoroughbred horses, using the primers # 1214 and # 1215. This PCR fragment contains a CT diallelic polymorphism (JH261). Following the hybridization of the single-stranded PCR templates to the GBA primer, an extension reaction was carried out using fluorescein-labeled ddCTP and biotin-labeled ddUTP. The two haptens were detected as described in the text. A, detection of ddCTP incorporation using an alkaline phosphatase conjugate; B, detection of ddUTP incorporation using a horseradish peroxidase conjugate. The plate contained the following controls: a) no template was added to wells A12 and B12 (template-independent extension controls); b) a 35 mer synthetic template (250 fmol) giving incorporation of a ddCTP was added to wells C12 and D12; c) a similar synthetic template giving incorporation of a ddUTP was added to wells E12 and F12; d) a mixture of both templates was added to wells G12 and H12; e) to control for PCR crosscontamination, negative PCR reactions were carried out and added to wells G11 and H11.

measured in a 96-well spectrophotometer. The graph in Figure 3 depicts the results quantitatively as a scatter plot. The values for each horse by typing for allele 1 (C) are indicated on the X-axis; those for allele 2 (T) on the Y-axis. Previous experiments which typed this locus with respect to several hundred horses failed to find a third allele (data not shown). The data in Figure 3 are consistent with JH261-1 being a diallelic, single nucleotide polymorphism (SNP).

The genotype groups are circled in Figure 3. A summary of the quantitative data is given in Table 2. It can be seen that the mean values for the two alleles can be calibrated to be roughly equivalent. Furthermore, variability between horses is surprisingly small. In the experiment shown, eight test samples produced signals with both alleles that were judged unscorable. Theoretically, this result could have been produced because of a failure in one or more of the biochemical reactions leading to the colorimetric data, because of a failure to amplify due to allelic variability in the primer sites, because of an allelic variability in the GBA primer site, or because the horses in question possessed an allele other than A or G in the template strand. We have investigated a large number of these results further and in all cases thus far examined, biochemical failure is the explanation of failure. This has been shown by analysis of the PCR reactions by gel electrophoresis to be PCR failure in most cases. The failure rate has been found to be substantially lower when a standardized control horse genomic DNA is used. For this reason, we believe that variability in sample quality produces most failure in our system. However, it can be anticipated that situations leading to failures of the other three types will arise in complex genomes and therefore adequate characterization of variability in target genetic loci is required for optimal utilization of the GBA method.

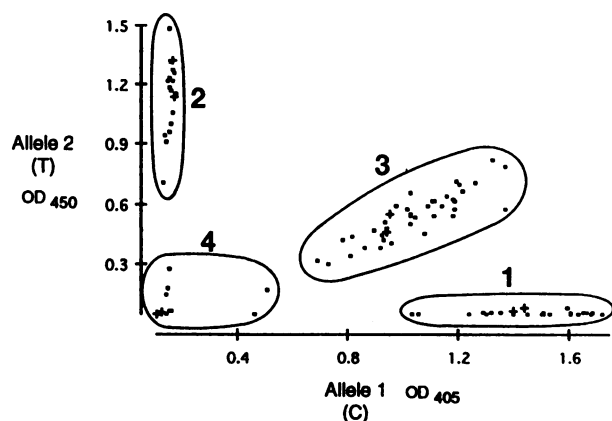


Figure 3. Scatter plot representation of the results from typing of 88 different thoroughbred horses with respect to the diallelic polymorphic locus JH261-1. The results shown are those from endpoint readings of the plate taken after 24 min of incubation with the enzyme substrates for the two antibody conjugates. Horse samples are indicated by dots while the control wells are plotted with '+'s. The points fall into four categories: 1) high values for C and low values for T; 2) high values for T and low values for C; 3) high values for both alleles; 4) high values for neither allele. The controls run in duplicate were: a) synthetic oligonucleotide molecules which mimic the PCR template for this locus and which possess a G at the variable position (+s found in group #1); b) which possess an A at the variable position (+s found in group #2); c) a mixture of these synthetic oligonucleotide templates (+s found in group #3); and d) PCR reactions to which no horse genomic DNA had been added as the amplification template (+s found in group #4).

Immobilization of the GBA primer to microtiter plates

We have previously described the details of our method for oligonucleotide immobilization onto polystyrene plates (25). Briefly, the method consists of incubating the oligonucleotide on the microtiter plates in a dilute solution of an organic or inorganic salt, followed by washing with a solution containing 0.05% Tween 20. We have shown that the oligonucleotides immobilized in this way are capable of specific hybridization to complementary templates and have used these findings to develop a convenient microplate-based PCR product detection assay. Other authors have also successfully immobilized oligonucleotide probes to the surface of polystyrene plates using NaCl-containing buffers and used those in hybridization-based assays (26).

In the current experiments, we have tested different compounds in the immobilization process and found that a number of chemically divergent reagents are capable of promoting this process. For example, the efficiency of immobilization of oligonucleotide #1112 using cetyltrimethylammonium bromide (CTAB) and tetramethylammonium chloride (TMAC) was compared (Figure 4). As negative controls, the oligonucleotides were added to some of the wells as aqueous solutions, without any immobilization reagents. To assess the immobilization process, following an overnight incubation with the immobilization reagents, the biotinylated oligonucleotide #1676, which is complementary to #1112, was added to the wells of the microtiter plate at a range of different concentrations. The amount of this biotinylated probe captured by hybridization to the immobilized oligonucleotide #1112 was determined by an enzyme-linked assay for the biotin residue. The results of this experiment are represented graphically in Fig. 4.

The results shown in Fig. 4 as well as similar results obtained with other compounds suggest that the immobilization reagents can be divided in two groups. The first group consists of chemicals like NaCl and TMAC, which work best when used at relatively high concentrations, generally higher than 50 mM, and best at 250 to 500 mM. Even concentrations as high as 1 M can be used without any noticeable adverse effect on the immobilization. The second group of immobilization reagents consists of chemicals that are characterized by the presence of two structural features: a positively charged 'head' and a relatively hydrophobic 'tail'. These are the typical features of cationic detergents. Representatives of this group are the cationic detergent cetyltrimethyl ammonium bromide (CTAB), octyldimethylamine hydrochloride, and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC). These compounds can be used for oligonucleotide immobilization at very low concentrations, as low as 0.03 mM for CTAB, but a lower hybridization signal is obtained when they are used at higher concentrations. The

Table 2. Typing of a single nucleotide polymorphism in equine DNA (locus JH261-1)

Genotype	Value AP	(SD)	Value HRP	(SD)
CC	1.478	0.190	0.051	0.009
CT	1.044	0.165	0.546	0.121
TT	0.155	0.012	1.143	0.186
NS	0.194	0.140	0.097	0.075

Endpoint readings were taken after 24 min of incubation with the colorimetric substrates. The average signals (in OD units at 405 and 450 nm) and corresponding standard deviations (SD) for 88 different horses are shown. (NS, no signal).

inhibitory concentrations differ among the reagents of this group. For CTAB, it is as low as 0.5 mM, whereas for EDC it is about 500 mM. It should be noted that the critical micelle concentration, CMC, for CTAB is about 1 mM. Thus, it is possible that once micelles are formed, the immobilization is inhibited. Compounds of a similar structure, but with a negatively charged 'head' (e.g., SDS) are completely inactive as oligonucleotide immobilization reagents, as are nonionic detergents.

The mechanism of immobilization in the presence of EDC or cationic detergents is probably very similar to the mechanism of transfer of nucleic acids and proteins through an organic phase in the presence of detergents described recently (27).

In another experiment, radioactively labeled oligonucleotides were immobilized to polystyrene plates using EDC. The amount of immobilized oligonucleotide was then determined by counting the amount of radioactivity released upon dissolving the wells in toluene. It was thus determined that approximately 1 pmole of oligonucleotide is immobilized in each well, which corresponds to about 10% of the input oligonucleotide (10 pmole). We also found that the input of oligonucleotide can be reduced to about 3 pmole per well before there is a noticeable decrease in the amount of immobilized material.

PCR amplification, generation of single-stranded DNA templates and their capture by hybridization to the GBA primers

PCR normally produces double-stranded products which do not hybridize to the immobilized capture oligonucleotide without prior strand separation. This strand separation can be achieved by treatment with heat or alkali, but we found the efficiency of hybridization to be low even with such a denaturation step. Asymmetric PCR has also been used for the generation of single-stranded products (28). Unfortunately, asymmetric PCR generates single-stranded products only linearly, and we found the results to be variable. Previously, we have reported a new and efficient method for the generation of single-stranded PCR products following a regular exponential amplification (25). The method is based on the selective protection of one of the strands of the PCR product from enzymatic hydrolysis of T7 gene 6 exonuclease

by the incorporation of four phosphorothioate bonds into the 5' end of that strand using modified PCR primers. The exonuclease method generates single-stranded products with high efficiency, and they are ideally suited for the subsequent hybridization to the immobilized primer.

The optimal length of the GBA primers immobilized in the microtiter plates appears to be between 20 and 25 bases. Oligonucleotides shorter than 20 bases usually give lower signals, and virtually no signals are seen with primers shorter than 10 bases. Apparently, parts of the immobilized oligonucleotides are inaccessible for hybridization because they are involved with interactions with the solid phase. This is supported by the finding that some 'hybrid' 25 mer primers that contain only about 12 to 15 bases at the 3' end exactly matching the template give signals as strong as those seen with completely matching 25 mers. Primers longer than 30 bases produce only slightly better extension signals, but tend to be more prone to template-independent extension (see below).

The solid phase primer extension reaction

In the enzymatic primer extension step, a single dideoxynucleoside triphosphate is incorporated at the 3' end of the immobilized GBA primer. The nature of the nucleotide at the polymorphic site of the template determines which of the four ddNTPs contained in the extension mixture will be incorporated by the polymerase. We have found that both the modified T7 DNA polymerase (Sequenase) and the Klenow fragment of *E. coli* DNA polymerase I are suitable enzymes for the primer extension. Both polymerases assure a very high signal-to-noise ratio. The Klenow polymerase possesses a 3'–5' exonucleolytic activity which Sequenase lacks (29). This exonuclease activity did not cause problems in the template-directed extension, but we have encountered cases of template-independent extension that are the result of this activity (see below).

During the development of GBA, we have found that false positive results can be generated by three different mechanisms. The first of these is trivial, and consists of self-extension at the 3' end of the template DNA at the same time as the 3' end of the immobilized GBA primer is being extended. This source of

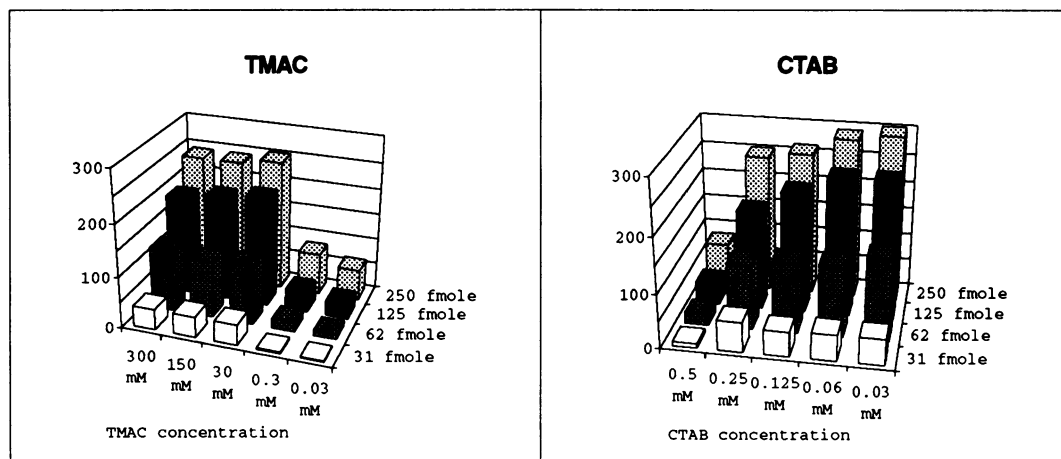


Figure 4. Comparison of CTAB and TMAC as oligonucleotide immobilization reagents. Oligonucleotide # 1112 was immobilized to Immulon 4 plates using varying concentrations of these two reagents, and then hybridized to the complementary biotinylated oligonucleotide # 1676, used at concentrations of 31, 62, 125, and 250 fmole per well. The signals obtained in the colorimetric assay are given in mOD₄₅₀/min.

'noise' is eliminated simply by briefly washing the plates after the polymerase extension step with a 0.2 N NaOH solution.

The second source of false positive signal, template-independent extension of the GBA primers, is the most likely problem to be encountered during the development of a GBA for a new polymorphism. This is the result of the formation of inter- or intramolecular secondary structures by the immobilized GBA primers and their subsequent enzymatic extension. Table 3 summarizes the results seen in the template-independent extension of two GBA primers, #308 and #680, as well as some of their modified versions.

A typical example of a primer showing template-independent extension is oligonucleotide #308. In this experiment, four separate wells were used to characterize the extension reaction. In each, only one of the ddNTPs was labeled with biotin while the other three were unlabeled. When this primer was immobilized on a plate, the extension reaction produced a strong signal due to incorporation of C. Analysis of the sequence of this oligonucleotide shows that it might be able to form relatively stable inter- or intramolecular partially self-complementary structures. These are shown in Figure 5. In both structures, the highlighted G residue will dictate the incorporation of a C by the polymerase.

To test whether these structures could explain the template-independent signal, a modified version of this primer was synthesized where the G residue of the original sequence was replaced by a T. This modified primer, #308T showed a strong template-independent extension signal in A. In another modified version of the same oligonucleotide, #308M1, the G residue of the original sequence was replaced by the abasic C₃ linker

Table 3. Template-independent extension of primers # 308, 680, and their modified versions

primer #	base G	base A	base T	base C
308	0.2	1.3	0.1	21.5
308T	2.0	48.5	0.2	0.2
308M1	0.5	0.3	0.9	0.5
680	0.4	1.2	0.5	65.5
680T	0.2	58.0	0.7	1.0

Extension reactions were done with the Klenow polymerase. The signals for the four different bases are given in mOD₄₅₀/min.

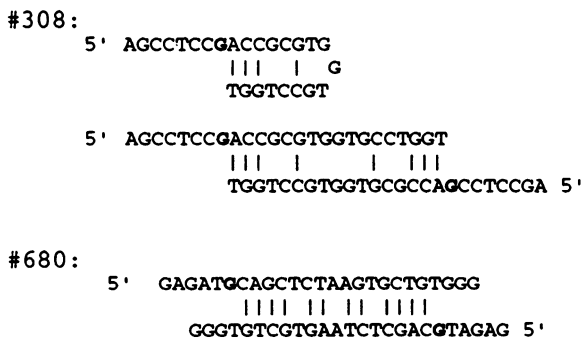


Figure 5. Postulated secondary structures of two GBA primers that could lead to the observed template-independent extensions. The nature of the incorporated nucleotides will be determined by the highlighted bases.

OPO₃CH₂CH₂CH₂ using a commercially available phosphoramidite. The modified primer #308M1 did not show any template-independent extension. The presence of the abasic linker within the sequence of 308M1 did not affect its hybridization and extension to synthetic or PCR-derived templates.

Primer #680 is an example of template-independent extension that is dependent on the polymerase used for extension. With this primer, a strong template-independent extension by a C was seen only when the extension was carried out with the Klenow polymerase (see Table 3), but not when the extension was carried out with the modified T7 DNA polymerase (Sequenase). It is reasonable to assume that this oligonucleotide forms the self-complementary structure shown in Figure 5. The four non-base paired residues are cleaved off by the 3'-5' exonuclease of the Klenow polymerase which then inserts a C opposite the highlighted G. When we replaced this deoxyguanosine residue by a thymidine in the modified primer #680T, the template-dependent noise was changed to an A, as expected (Table 3). Because of the possibility of template-independent extension of the GBA primers, each new GBA primer should be tested for this type of extension before being used in typing experiments. This phenomenon would produce inappropriate typings in some percentage of cases especially when PCR yield has been low. Commercially available computer programs for DNA analysis (e.g., 'Oligo', National Biosciences, Inc., Plymouth, MN) can be used to predict potential secondary structures. The required modifications can be incorporated in the sequence of the appropriate GBA primer.

The third source of false positive signals is template-dependent, but in contrast to the first type of 'noise' described above, it is the result of extension at the 3' end of the GBA primer and not the template. Signals of variable strength are sometimes generated in one of the 'wrong' bases, i.e., a base that is not consistent with the sequence of the template to be typed. We have observed the same 'noise' profile for both polymerases tested, Sequenase and Klenow. This type of 'noise' is notably dependent on the amount of template molecules hybridized to the GBA primer and can be especially serious when high concentrations (usually, more than 500 fmole) of synthetic template molecules are used in GBA. In the majority of cases when PCR generated templates are typed, this type of noise is undetectable or very weak (signal-to-noise ratios: >20), but on rare occasions can be strong enough to cause false interpretation of the genotyping results. A summary of the results of some GBA experiments with template-dependent 'noise' is given in Table 4.

The biochemical basis of this type of 'noise' is uncertain, but it may be the result of mishybridization of the GBA primer to

Table 4. Analysis of template-dependent noise

GBA experiment	base G	base A	base T	base C
713+1302	12.0	35.1	105.4	35.5
713+1473	60.4	50.8	160.0	2.5
713+1474	23.4	64.7	120.4	2.5
501+1464	5.5	12.4	85.0	1.0
(in solution)				
501+1464	7.0	35.5	170.2	2.0
(solid phase)				

Signals are given in mOD₄₅₀/min. 500 fmole of the synthetic templates was used in the solid phase extension experiments involving primer #713. Incorporation of a labeled T was expected in all of these experiments.

the template and/or misincorporation of the wrong ddNTP by the polymerase. Indeed, the polymerases used may in some sequence contexts display a higher misincorporation rate with the labeled ddNTPs used. We have analyzed in more detail the template-dependent 'noise' observed with the GBA primer #713 and the synthetic template #1302 (see Table 4). In GBA, this template-primer combination gives significant template-dependent 'noise' in bases C and in A. We synthesized and tested two modified templates, #1473 and #1474, which differ from #1302 only by a few bases at the 5' end, i.e., in a part of the template that is not expected to hybridize to the GBA primer. In oligonucleotide #1473, three deoxyguanosine residues of #1302 are changed to deoxycytidines; in oligonucleotide #1474, the part of the synthetic template extending beyond the GBA primer is reduced to one single deoxyadenosine. The results shown in Table 4 demonstrate that the template-dependent 'noise' is influenced by the sequence surrounding the residue of the template DNA that directs the dideoxynucleotide incorporation. Thus, the replacement of the deoxyguanosines of #1302 with deoxycytidines in #1473, or their elimination in #1474, reduces the 'noise' in C and, in the case of template #1473, increases the noise in G. Such context-dependent effects on the fidelity of DNA polymerases have been reported before (30).

To verify that this 'noise' is not due to the fact that the hybridization and extension reactions are carried out on the solid phase, an extension experiment was carried out in solution (see Experimental). The 5' biotinylated oligonucleotide #1464 was used as a primer and annealed to the oligonucleotide #501. The expected signal in this template-primer combination is a T. In parallel, this typing experiment was carried out on the polystyrene solid phase, by immobilizing the primer #501 and using 500 fmole of oligonucleotide #1464 as the template. These two experiments produced a remarkably similar signal-to-noise profiles (see Table 4). Analogous results were obtained with other primer-template combinations (not shown).

Although we have not found a general solution for eliminating the rare occurrence of template-dependent 'noise', the following approaches have been found to reduce or eliminate the problem. Hybrid GBA primers of the type 5' X₁₂N₁₃, where each X position contains equal amounts of the four bases whereas the 13 bases at the 3' end match exactly the template were sometimes found to give better signal-to-noise ratios than completely matching 25 mers. This could be due to reduced or eliminated mishybridization with these primers. The signal-to-noise ratio was also improved by performing the extension reaction at 5°C rather than at room temperature, and also by decreasing the concentration of all ddNTPs in this step. These two factors probably affect the fidelity of the polymerase. Finally, this 'noise' can usually be avoided by switching the primer protected from exonuclease digestion and typing the same polymorphism on the opposite DNA strand with a suitable GBA primer.

Colorimetric detection of the incorporated labeled nucleotide

As shown above, it is possible to type single nucleotide polymorphisms by GBA by including only one labeled ddNTP per well. However, the use of two labeled ddNTPs allows the determination of both alleles in a diallelic locus to be carried out in the same well. This not only reduces the amount of PCR generated template required and results in considerable savings of labeled chain terminators, but serves as a very useful internal control for all post-PCR steps. For example, for a particular

template, a blank well in the one-base-per-well mode could be due to homozygosity of the other type but also to failure of one of the post-PCR steps (hybridization, extension). In the two-bases-per-well mode, lack of signal for one of the bases can only be due to homozygosity of the other type or failure of the enzyme-linked assay for this allele. The latter hypothesis can be routinely excluded with suitable controls.

To date, we have found no randomly discovered site of single nucleotide polymorphism to be tri- or tetra-allelic. This is consistent with findings in other laboratories (Deborah Nickerson, personal communication). However, if such a site were to be encountered, it could be typed using a second well whereby incorporation of the other two nucleotides could be examined through inclusion of their ddNTP derivatives. Alternatively, other haptenated ddNTPs could be used simultaneously assuming an appropriate antibody-enzyme conjugate was available. Failure to test for the alleles which are rare in the population can produce incompatible data in legitimate pedigrees because heterozygotes where one chromosome possesses a 'null' allele would appear to be a homozygote for one of the tested alleles. A falsely typed heterozygous parent would appear to be excluded if its offspring inherits the 'null' allele. This explanation can be considered likely if many single-locus exclusions are observed for a particular locus in a panel of markers used for genetic studies.

CONCLUSION

Compelling arguments exist for the development of DNA-based assays, which can be performed on a very large scale, for the analysis of known polymorphisms in complex genomes. In this paper we give biochemical details concerning a new genotyping procedure, GBA (genetic bit analysis), that is simple, convenient, and automatable. In this method, sequence-specific primer annealing is used to select a unique polymorphic site in a nucleic acid sample, and interrogation of this site is accomplished via the highly accurate DNA polymerase reaction using a set of novel, commercially available, non-radioactive dideoxynucleotide analogs.

An important feature of GBA is that it does not rely on nucleic acid hybridization for purposes of nucleotide discrimination. This is in marked contrast to methods such as allele-specific hybridization (12). Rather, nucleic acid hybridization is used only to position the template to be typed with respect to a primer molecule. Nucleotide discrimination is accomplished using a DNA polymerase, an enzyme that has evolved to perform that role. As a result, all GBA reactions can be performed under a standard set of conditions that have been optimized for high throughput operations. Many loci can be tested simultaneously, and new tests can be developed rapidly.

GBA was developed to be a method that can be applied on a very large scale using commercial liquid handling devices. In our parentage verification assay for thoroughbred horses, as many as 180 horses are typed by GBA at 26 diallelic loci by a single technician in a single day. However, GBA can easily be carried out manually. The signals generated are usually strong enough to allow visual interpretation of the results (Figure 2), without the need for a spectrophotometer. Alternatively, data can be acquired on a large scale using high-volume, stacking plate readers. Therefore, it has the potential to become a method for the typing of single nucleotide polymorphisms across a broad spectrum of applications.

ACKNOWLEDGEMENTS

Special thanks are due to Dr Jonathan Karn for his advice and continuous support of this project. We are grateful to Dr Sidney Brenner for stimulating discussions. The technical assistance of Dr Florence McAuliffe and Mr Richard Douglas in the early stages of this work is also gratefully acknowledged.

REFERENCES

- Cooper, D.N., Smith, B.A., Cooke, H.J., Niemann, S. and Schmidtke, J. (1985) *Hum. Genet.* 69, 201–205.
- Kerem, B.-S., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M. and Tsui, L.-C. (1989) *Science* 245, 1073–1080.
- Watson, J.T., Marotta, C.A., Forget, B.G. and Weissman, S.M. (1977) *Trans. Assoc. Am. Physicians* 90, 117–126.
- Fearon, E.R. and Vogelstein, B. (1990) *Cell* 61, 759–767.
- Sung, C.H., Davenport, C.M., Hennessey, J.C., Maumenee, I.H., Jacobson, S.G., Heckenlively, J.R., Nowakowski, R., Fishman, G., Gouras, P. and Nathans, J. (1991) *Proc. Natl. Acad. Sci. USA* 88, 6481–6485.
- Dryja, T.P., Hahn, L.B., Cowley, G.S., McGee, T.L. and Berson, E.L. (1991) *Proc. Natl. Acad. Sci. USA* 88, 9370–9374.
- Botstein, D., White, R.L., Skolnick, M. and Davis, R.W. (1980) *Am. J. Hum. Genet.* 32, 314–331.
- Myers, R.M., Maniatis, T. and Lerman, L.S., (1987) *Methods Enzymol.* 155, 501–527.
- Hayashi, K. (1991) *PCR Methods Applic.* 1, 34–38.
- Meyers, R.M., Larin, Z. and Maniatis, T. (1985) *Science* 230, 1242–1246.
- Cotton, R.G.H., Rodrigues, N.R. and Campbell, R.D. (1988) *Proc. Natl. Acad. Sci. USA* 85, 4397–4401.
- Conner, B.J., Reyes, A.A., Morin, C., Itakura, K., Teplitz, R.L. and Wallace, R.B. (1983) *Proc. Natl. Acad. Sci. USA* 80, 278–282.
- Saiki, R.K., Walsh, P.S., Levenson, C.H. and Erlich, H.A. (1989) *Proc. Natl. Acad. Sci. USA* 86, 6230–6234.
- Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) *Science*, 241, 1077–1080.
- Nickerson, D.A., Kaiser, R., Lappin, S., Stewart, J., Hood, L. and Landegren, U. (1990) *Proc. Natl. Acad. Sci. USA* 87, 8923–8927.
- Barany, F. (1991) *Proc. Natl. Acad. Sci. USA* 88, 189–193.
- Wu, D.Y., Ugozzoli, L., Pal, B.K. and Wallace, R.B. (1989) *Proc. Natl. Acad. Sci. USA* 86, 2757–2760.
- Newton, C.R., Graham, A., Heptinstall, L. E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J. C. and Markham, A.F. (1989) *Nucleic Acids Res.* 17, 2503–2516.
- Syvänen, A.-C., Aalto-Setälä, K., Harju, L., Kontula, K. and Soderlund, H. (1990) *Genomics* 8, 684–692.
- Kuppuswami, M.N., Hoffmann, J.W., Kasper, C.K., Spitzer, S.G., Groce, S.L. and Bajaj, S.P. (1991) *Proc. Natl. Acad. Sci. USA* 88, 1143–1147.
- Nyren, P., Pettersson, B. and Uhlen, M. (1993) *Anal. Biochem.* 208, 171–175.
- Ugozzoli, L., Wahlquist, J.M., Ehsani, A., Kaplan, B.E. and Wallace, R.B. (1992) *GATA* 9, 107–112.
- Syvänen, A.-C., Sajantila, A. and Lukka, M. (1993) *Am. J. Hum. Genet.* 52, 46–59.
- Prober, J.M., Trainor, G.L., Dam, R.D., Hobbs, F.W., Robertson, C.W., Zagurski, R.J., Cocuzza, A.J., Jensen, M. and Baumeister, K. (1987) *Science* 238, 336–341.
- Nikiforov, T.T., Rendle, R.B., Kotewicz, M.L. and Rogers, Y.-H. (1994) *PCR Methods Applic.* 3, 285–291.
- Balaguer, P., Térouanne, B., Allibert, P., Cros, P., Boussioux, A.-M., Mandrand, B. and Nicolas, J.-C. (1993) *Mol. Cell. Probes* 7, 155–159.
- Bromberg, L.E. and Klivanov, A.M. (1994) *Proc. Natl. Acad. Sci. USA* 91, 143–147.
- Gyllenstein, U. and Erlich, H.A. (1988) *Proc. Natl. Acad. Sci. USA* 85, 7652–7656.
- Tabor, S. and Richardson, CC. (1989) *J. Biol. Chem.* 264, 6447–6458.
- Goodman, M.F., Creighton, S., Bloom, L.B. and Petruska, J. (1993) *Crit. Rev. Biochem. Mol. Biol.* 28, 83–126.